Regresión Lineal Simple

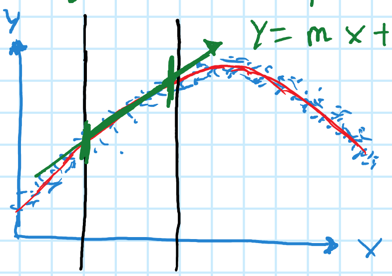
jueves, 2 de mayo de 2019 04:28 p. m.

- Controlar la variable independiente
- Observar el comportamiento de la variable independiente

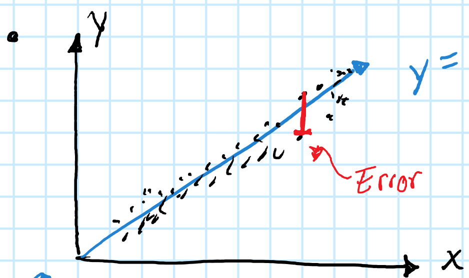
// Lineal en los coeficientes no en la variable

// No podemos usar el modelo obtenido en diferentes secciones

Diagrama de Dispersión



- Ajuste de la recta de regresión por de métodos de mínimos cuadrados ← De los errores

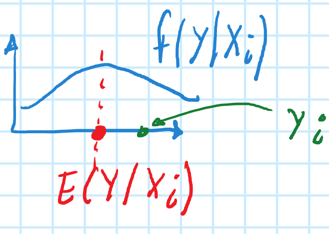


$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x$ / Estimador

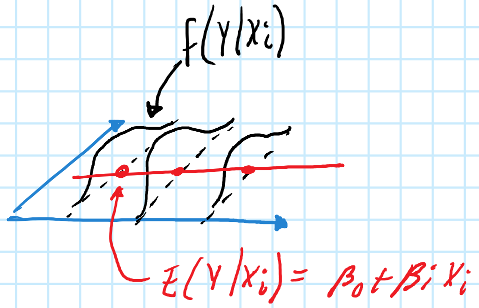
- Prueba de hipótesis sobre la pendiente de la recta

- Los supuestos del modelo de regresión lineal simple son que los errores ε_i se distribuyen de manera normal con media cero y varianza σ^2

En un punto



En la recta



$$\varepsilon \sim N(0, \sigma^2)$$

$\varepsilon \sim N(0, \sigma^2)$

// Mínimos Cuadrados ⋮ $\sum_{i=1}^{n} \varepsilon_i^2$ sea mínima

$$Y_i = \beta_0 + \beta_i x_i + \varepsilon_i$$

$$\varepsilon_i = Y_i - \beta_0 - \beta_i x \xrightarrow[\text{al cuadrado}]{\text{Elevado}} \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_i x)^2$$

Simplificación de la expresión

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$$

$$\hat{Y} = \beta_0 + \beta_1 x$$

$$\beta_1 = \frac{\sum_{i=1}^{n} x_i Y_i - \frac{\left(\sum_{i=1}^{n} Y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}$$

covarianza

$$\beta_1 = \frac{\sum_{i=1}^{n} x_i Y_i - \frac{\left(\sum_{i=1}^{n} Y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \Rightarrow \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

- $S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = \sum_{i=1}^{n} (x_i - \bar{x})^2$
- $S_{xy} = \sum_{i=1}^{n} x_i Y_i - \frac{\left(\sum_{i=1}^{n} Y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n} = \sum_{i=1}^{n} (x_i - \bar{x}) Y_i$

Resumen! $S_{xx} = \sum (x_i - \bar{x})^2$

Resumen

$S_{xx} = \Sigma (x_i - \bar{x})^2$

$S_{yy} = \Sigma (y_i - \bar{y})^2$

$S_{xy} = \Sigma (x_i - \bar{x})^2 y_i = \Sigma x_i y_i - n\bar{y}\bar{x}$

// Recta Estimada de Regresión

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

- $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-1.028}{79.784} \cong -.0129$

- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 24.80 - (-.0129)(24.88) =$

$= 25.1272$

$\therefore \hat{y} = 25.1272 - .0129x$

Conclusiones: x prácticamente no afecta al comportamiento de $\hat{y}$ lo que significa que puede existir otro fenómeno que afecte

// No hay relación lineal entre "x" y "y"

Coeficientes

martes, 14 de mayo de 2019 09:09 a. m.

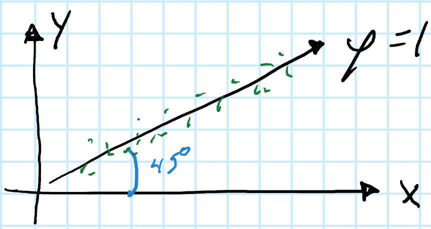
- Ajuste de la recta de regresión por el método de mínimos cuadrados
- Pruebas de hipótesis sobre los parámetros de regresión

"$H_0: \beta_0 = 0$" o "$H_1: \beta_1 \neq 0$" | β_0 y β_1 parámetros

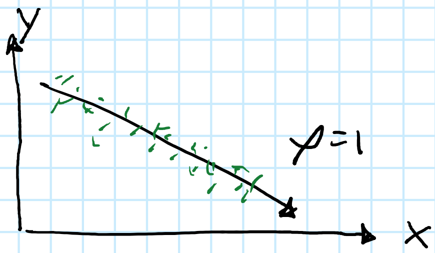
- Coeficientes de correlación y determinación y su interpretación

Coeficiente de Correlación de 2 variables conjuntas X y Y | $\bullet\ \gamma = \dfrac{\sigma_{xy}}{\sigma_{xx}\,\sigma_{YY}}$

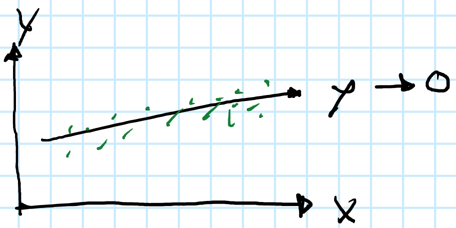
→ Si existe una relación funcional entre las 2 variables. (lineal)



Correlación perfecta y una gran probabilidad de que x y Y están asociados



Correlación perfecta y una gran probabilidad de que "X" y "Y" están asociados
// Valores pequeños de "Y" o "X"



No hay correlación entre "X" y "Y"

// El modelo de regresión no debe asumirse como una

// El modelo de regresión no debe asumirse como una relación causa-efecto, sino únicamente como una variación conjunta

El estimador insesgado del Coeficiente de Correlación ; $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

Coeficiente de Determinación ; $R^2 = r^2$

→ Indica la proporción (o porcentaje) de la variación "Y" que se debe a la variación de "X"

→ Que tanto de la variación de "Y" es explicada por la variable independiente

- Para realizar la significación de la regresión se pueden hacer las pruebas

// Para la pendiente

$H_0: \beta_1 = 0$ | $H_0: \rho = 1$

$H_1: \beta_1 \neq 0$ | $H_1: \rho \neq 1$

- También se pueden construir intervalos de confianza

a) Para la respuesta Y dada un X_0

b) Para la pendiente β_0

c) Para ρ

- El estimador es $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ donde: $\hat{\beta}_1 \sim$ Normal

y $E\{\hat{\beta}_1\} = \hat{\beta}$

$V\{\hat{\beta}_1\} = \frac{\sigma^2}{S_{xx}}$ ← Varianza de los errores

// σ^2 es desconocida se usa su estimador

$\therefore \hat{V} = \frac{MSE}{S_{xx}}$ ← Cuadrado medio de los errores

y la V.A

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{MSE}{S_{xx}}}}$$

// Tiene n-2 grados de libertad

// Realizando Hipótesis

→ Estadístico de Prueba

- $T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{MSE}{S_{xx}}}}$ de nivel de significancia 5 %

- $\frac{\alpha}{2} = .025$

$t_{.025,\,13} = 2.16$ ($13 = n-2$)

// Prueba de 2 colas

$RR = \{T_m < -2.16 \cup T_m > 2.16\}$

// Sustituyendo en expresión

Hipótesis $\beta_1 = 0$

$T = \frac{-.012 - 0}{\sqrt{\ldots}}$

$$T = \frac{-.012 - 0}{\sqrt{\frac{5.2396}{79.784}}} \leftarrow MSE = \frac{SS_E}{n-2}$$

$$= -0.052$$

$$SS_E = S_{YY} - \hat{\beta}_1 S_{XY}$$

Conclusión: Al nivel de significancia del 5% (No está entre los intervalos) no hay evidencias suficientes que indique que la pendiente es diferente de cero.

→ La resistencia al corte no depende de la tensión del material.

// Intervalos de Confianza

$$\beta_1 \in \left(\hat{\beta}_1 \pm t_{\frac{\alpha}{2}, n-2} \sqrt{\frac{MSE}{S_{XX}}} \right)$$

$$\beta_1 \in \left(-1.5 \pm (-3.1824)\left(\sqrt{\frac{1/3}{4}}\right)\right)$$

$$-2.48 < \beta_1 < -.58$$